

# Linguistic Markers of Status in Food Culture: Bourdieu's Distinction in a Menu Corpus

Dan Jurafsky, Victor Chahuneau, Bryan R. Routledge, Noah A. Smith

10.18.16

Peer-Reviewed By: Nathalie Cooke & Anon.

Clusters: Food

Article DOI: 10.22148/16.007

Dataverse DOI: 10.7910/DVN/QMLCPD

Food is a core element of culture, whose link with identity and socio-economic class has made it an important area of cultural research.<sup>1</sup> In his ground-breaking study, Pierre Bourdieu noted that “oppositions similar in structure to those found in cultural practices also appear in eating habits.”<sup>2</sup> His work established deep associations linking food culture, and taste more generally, with social class and other aspects of identity, demonstrating the economic and social determinants of taste and their role in representing *distinctions*, differences between groups.

These implicit assumptions about taste can nonetheless be difficult to tease out from social investigations, such as the careful survey analysis or explicit interviews that Bourdieu conducted in the 1960s. However, there is a more direct window onto implicit assumptions about culture: the language we use in talking about culture. Language offers a powerful tool for observing and quantifying the sometimes unconscious way that our associations and understandings of culture reflect our social attitudes and prejudices. Studying social aspects of culture, however, requires controlling for the great number of confounds that come along with language. It's hugely helpful therefore to focus on a single genre, allowing exploration of a single contextual domain of language, and to use sufficiently powerful data that digital methods can be used to control for other confounds.

We propose to study the reflections of Bourdieu's distinction in the language used on menus of restaurants in the United States. By focusing on menus, we remove the many possible confounds due to genre: menus offer a coherent context, one in which food producers are attempting to advertise their food to consumers, framing it in a way that reflects how they believe their consumers understand food. Menus cross socio-economic lines, offering text written by restaurants from the cheapest and most quotidian to the most expensive, luxurious, and high-status, and offering very explicit coding of price. Finally, menus are frequent, allowing us to build a corpus of menus large enough that it is possible to control for cultural effects like different ethnicities of restaurants, geographical effects like city or neighborhood, and linguistic effects like variation in numbers of words and of dishes.

A wide variety of research has drawn on the insights of Bourdieu to examine the links between, status, class, and culture on food. Building on this literature, we focus on four aspects in which the differences between the language used by inexpensive restaurants versus expensive restaurants on their menus reflects a particular aspect of the framing of distinction.

The first is the role of *authenticity*. A number of scholars have studied how expensive or high-status food is portrayed as authentic.<sup>3</sup> Beverland and colleagues interviewed luxury winemakers and consumers, and consumers of Trappist beer

<sup>1</sup>This work was supported in part by a research grant from Google, and by the Center for Advanced Study in the Behavioral Sciences at Stanford University. We are grateful for very helpful suggestions from David Bamman, Mark Liberman, and the anonymous reviewers, and to the New York Public Library for their generosity in making the Buttolph menu collection available to the public.

<sup>2</sup>Pierre Bourdieu, *Distinction: A Social Critique of the Judgement of Taste*, trans. Richard Nice (Harvard University Press, 1984).

<sup>3</sup>Glenn R Carroll and Dennis Ray Wheaton. “The organizational construction of authenticity: An examination of contemporary food and dining in the U.S.” *Research in Organizational Behavior* 29 (2009): 255-282.

brewed in Belgium and the Netherlands. Both consumers and producers link authenticity with historicity (talking about the founding of the company), with the relationship to place (in particular the concept of *terroir*) and with a focus on traditional, hand-crafted methods of production.<sup>4</sup> Johnston and Bauman looked at every article that appeared in 2004 in four upscale food magazines (*Bon Appétit*, *Saveur*, *Food and Wine*, and *Gourmet*), showing they present food as authentic by talking about locality and the regions in which food is produced, by the use of handmade simplicity rather than industrial production, and by the historicism indicated via a long tradition of manufacture.<sup>5</sup> Lakoff performed a careful discourse analysis comparing a menu from one expensive restaurant (*Chez Panisse*) with one from an inexpensive one nearby (*The Oriental Restaurant*), noting among other differences that the expensive menu contains extensive reference to the provenance of their food on particular farms. She argues that the references on the menu to local farms and ranches “that practice ecologically sound agriculture” allow eating a meal to be an act of ‘civic virtue’, suggesting that this kind of authenticity is also associated with a particular aspect of morality.<sup>6</sup>

While authenticity has mainly been considered as a property of high-status foods, recent work in our labs has also explored the role of authenticity in signaling less expensive foods. Freedman and Jurafsky investigated the advertising language on bags of potato chips and the link between language and the price of the chips. They found that advertising on expensive chips emphasizes what they called *natural authenticity*, by mentioning the provenance and quality of the ingredients (*sea salt*, *Yukon Gold potatoes*) and the hand processing of the ingredients (*hand-rake every batch*), and using the words *natural* and *organic*. Advertising on less expensive chips, by contrast, was consistent with a different model that they called *traditional authenticity*, in which food was related to family members, historicity, and family tradition (*old family recipe*, *time-honored tradition*). They suggested their results are consistent with earlier work suggesting that working class or lower-middle-class identity is likely to be based around family and tradition. Similarly, Chahuneau et al. suggested that inexpensive menus employ *traditional authenticity*, relating food to family, comfort, and old-fashioned tradition.<sup>7</sup>

We therefore propose to examine menus to see if they indeed present these two aspects of authenticity. Our hypothesis is that more expensive, higher-status restaurants will be more likely to emphasize natural authenticity and the provenance of the ingredients on the farm, while cheaper, lower-status restaurants will be more likely to emphasize traditional authenticity.

The second aspect of distinction we consider is *educational capital*. Since education is associated with socioeconomic status, expensive restaurants, like other advertisers of luxury products, may use “fancier” words, i.e. ones that are rarer, more morphologically complex, or drawn from high-status foreign languages. The goal of such language may be to signal status or target educated consumers, drawing consumers into believing that the product is consonant with their educational capital and flattering the consumer by expressing this shared knowledge.<sup>8</sup> In his 1963 *Confessions of an Advertising Man* David Ogilvie warns would-be ad-writers: “Don’t use highfalutin language” when you’re talking to a non-highfalutin audience. Freedman and Jurafsky found that advertising text on the bags of more expensive potato chips indeed tend to use words which are longer and more complex. Silverstein analyzed wine tasting notes, suggesting that aspirational elites self-consciously use rare or difficult wine-tasting jargon words (‘oinoglossia’) as an attempt by to demonstrate their prestige simply through the very act of using elite words.<sup>9</sup>

<sup>4</sup>Michael B. Beverland, “Crafting brand authenticity: The case of luxury wines,” *Journal of Management Studies*, 42, no. 5 (2005):1003-1029. Michael B. Beverland, “The ‘real thing’: Branding authenticity in the luxury wine trade,” *Journal of Business Research* 59, no. 2 (2006): 251-258. Michael B. Beverland, Adam Lindgreen, and Michiel W. Vink, “Projecting Authenticity Through Advertising: Consumer judgments of advertisers’ claims,” *Journal of Advertising* 37, no. 1 (2008): 5-15.

<sup>5</sup>Josée Johnston and Shyon Baumann, “Democracy versus Distinction: A Study of Omnivorousness in Gourmet Food Writing,” *American Journal of Sociology* 113, no. 1 (2007): 165-204.

<sup>6</sup>Robin Lakoff, “Identity a la Carte; or, You Are What You Eat,” In *Discourse and identity (Studies in interactional sociolinguistics)*, ed. Anna De Fina, Deborah Schiffrin, Michael Bamberg (New York: Cambridge University Press, 2006), 147-165.

<sup>7</sup>Our labs: Joshua Freedman and Dan Jurafsky, “Authenticity in America: Class distinctions in potato chip advertising,” *Gastronomica: The Journal of Critical Food Studies* 11, no. 4 (2011): 46-54; Victor Chahuneau, Kevin Gimpel, Bryan R. Routledge, Lily Scherlis, and Noah A. Smith, “Word salad: Relating food prices and descriptions,” In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, (Association for Computational Linguistics: 2012), 1357-1367. Re: working class identity see Douglas B. Holt, “Does cultural capital structure American consumption?” *Journal of Consumer Research* 25, no. 1 (1998): 1-25; Marjorie DeVault, *Feeding the Family: The Social Organization of Caring as Gendered Work* (University of Chicago Press, 1994); Bourdieu 1984; Joan C. Williams, *Reshaping the Work-Family Debate: Why Men and Class Matter* (Harvard University Press, 2010); Alana Conner Snibbe and Hazel Rose Markus, “You can’t always get what you want: educational attainment, agency, and choice,” *Journal of personality and social psychology* 88, no. 4 (2005): 703-720; Nicole M. Stephens, Stephanie A. Fryberg, and Hazel Rose Markus, “When Choice Does Not Equal Freedom: A Sociocultural Analysis of Agency in Working-Class American Contexts,” *Social Psychological and Personality Science* 2, no. 1 (2011): 33-41.

<sup>8</sup>David Ogilvy, *Confessions of an Advertising Man*; Michael Silverstein, “Indexical order and the dialectics of sociolinguistic life,” *Language and Communication* 23 (2003): 193-229; Lakoff “Identity a la Carte”; Freedman and Jurafsky “Authenticity in America”.

<sup>9</sup>Silverstein “Indexical order and the dialectics of sociolinguistic life”

Our hypothesis is that more expensive, higher-status restaurants will use more complex language, while cheaper, lower-status restaurants will use simpler language.

The third aspect of distinction we consider is Bourdieu's concept of *plenty* in the working-class meal. In his study of French society in the 1960s, Bourdieu noted:

The working-class meal is characterized by plenty... "Elastic" and "abundant" dishes are brought to the table—soups or sauces, pasta or potatoes (almost always included among the vegetables)—and served with a ladle or spoon, to avoid too much measuring and counting...<sup>10</sup>

Bourdieu suggests a framing for working-class meals in which the "impression of abundance" is important, and in which food should be simultaneously cheap and "nourishing" or "filling".

Does this same framing of plenty associated with cheaper meals apply to modern restaurant meals? In a related finding, Lakoff in her comparison of menus from an expensive restaurant (*Chez Panisse*) versus a cheap one (*The Oriental Restaurant*) noted that the cheaper restaurant gave the diner vastly more choices (with a longer menu and choices between chicken or pork or shrimp, for example).<sup>11</sup> Her finding suggests that inexpensive restaurants may focus on diner choice as an aspect of plenty: not only does the restaurant focus on offering more food, but also on offering many more kinds of food. We therefore hypothesize that menus of cheaper restaurants will be more likely to give linguistic indicators of plenty and choice—the menu will highlight the *amount* of food, offering more dishes and offering a greater range of choices, while menus of expensive restaurants will be less likely to use this framing.

If so, we would expect a linguistic framing of plenty emphasized more in cheaper restaurants than in expensive ones.

The fourth aspect of distinction we consider is "implicit signaling of quality": whether expensive restaurants are likely to be less explicit in signalling the quality of their products, while cheaper or middle-priced restaurants are more explicit. Bourdieu, as Ridgeway and Fisk point out, "comments ... on the insecurity of the (middle class) petit bourgeois and nouveau riche compared to the upper class 'who ... have the privilege of not worrying about their distinction.'" <sup>12</sup> Liberman suggests that this idea of middle class insecurity might explain the differences he found among menus of 3 restaurants (cheap, middle-priced, and expensive) in Philadelphia. He noted that the middle priced restaurant had the wordiest menu (with a hamburger described as "Big Juicy Burger of Buck Run Farm's Grass Fed Beef on our House made Poppy Seed Bun"). Liberman argued that the elaborate language in middle priced restaurant menus is used as an index of status, and that the extensive modifications by words like *juicy* or *peppery* is a mark of status anxiety.<sup>13</sup> Liberman's idea is supported by research in game-theoretic models of advertising which suggests that middle-priced firms will tend to use explicit advertising and that high-priced firms will tend to use implicit or "modest" advertising, "counter-signaling", to distinguish themselves from the middle-priced firms.<sup>14</sup>

We therefore hypothesize that explicit protestations of quality (adjectives like "real", "delicious", "crispy", "crunchy") will be more likely to appear in middle-priced or cheaper restaurants than in expensive ones. We focus on adjectives because a wide variety of researchers have noted that adjectives ("fresh", "crispy", "fluffy", "buttery") richly populate menus and that customers are more likely to choose dishes with these words.<sup>15</sup>

In the next sections we develop measures for these four aspects of distinction and introduce the corpus we use to evaluate them. In a later section we also introduce a historical corpus to investigate the diachronic nature of our hypothesis.

<sup>10</sup>Bourdieu *Distinction*, 194.

<sup>11</sup>Lakoff, "Identity a la Carte".

<sup>12</sup>Cecilia L. Ridgeway and Susan R. Fisk, "Class rules, status dynamics, and 'gateway' interactions," In *Facing Social Class: How Societal Rank Influences Interaction*, ed. Susan T. Fiske and Hazel Rose Markus (New York: Russell Sage Foundation, 2012.), 249.

<sup>13</sup>Mark Liberman, "Modification as social anxiety," *Language Log* (May 16, 2004).

<sup>14</sup>Nick Feltovich, Rick Harbaugh and Ted To, "Too cool for school? Signalling and countersignalling," *RAND Journal of Economics* 33, no. 4, (2002): 630-649; Ram Orzach, Per Baltzer Overgaard and Yair Tauman, "Modest advertising signals strength," *RAND Journal of Economics* 33, no. 2 (2002): 340-358.

<sup>15</sup>Ann Zwicky and Arnold Zwicky, "America's National Dish: The Style of Restaurant Menus," *American Speech* 55, no. 83 (1980): 87-92; Brian Wansink, Koert van Ittersum, and James E. Painter, "How Descriptive Food Names Bias Sensory Perceptions in Restaurants," *Food Quality and Preference* 16, no. 5 (2005): 393-400; Brian Wansink, James M. Painter, and Koert Van Ittersum, "Descriptive menu labels' effect on sales," *Cornell Hotel and Restaurant Administrative Quarterly*, 42 (2001), 68-72.

## The corpus

We chose a dataset of menus large enough to investigate the full range of restaurants, from fast food to luxury restaurants, allowing the investigation of the linguistic strategies of distinction jointly while controlling for each other and many potential confounding factors like the geographical location, the type of food served at the restaurant, the total number of words, and so on. We built our dataset by extending the corpus of Chahuneau et al. (2011), consisting of menus downloaded from the website *allmenus.com* in 2011 for restaurants in seven cities: Boston, Chicago, Los Angeles, New York, Philadelphia, San Francisco, and Washington D.C. They randomly divided the restaurants into 80 percent for training and 20 percent reserved for evaluation and testing. All the analyses in this paper are performed on their training dataset. From this data, we used only restaurants that were characterized on Yelp as restaurants and bars; thus all delis, groceries, and caterers were removed from the dataset.

Each menu comprises a set of dishes with names, descriptions, and a price. For each restaurant, we use two metadata variables drawn from the reviewing website *yelp.com*: the restaurant category (type of food), and the price range, a variable on a four-level scale from \$ to \$\$\$\$\$. We use this Yelp price range (\$, \$\$, \$\$\$, \$\$\$\$) as our measure of restaurant price class, considering \$\$\$\$ as upper-socioeconomic-status, \$\$ and \$\$\$ as middle-status, and \$ as lower status. The Yelp price range is assigned with respect to restaurant category (a pizza restaurant that is rated as \$\$\$ will be less expensive than a French restaurant rated \$\$\$), so we control for category whenever investigating the price range variable. Because all the data was automatically downloaded from the web, we also performed a number of error-checking and correction operations. Dozens of restaurants did not have restaurant category or price range variables on Yelp; these were coded by hand. Seven restaurant menus were removed from the study due to significant errors in price data (caused by typos due to download errors). We also removed all restaurants with missing menu data. Finally, because we focus only on food dishes, all drinks (sodas, alcohol, coffee) are removed from the lists of dishes, by using a combination of hand-labeling and regular expressions; in total 45,018 menu descriptions of drink items were removed.

The resulting dataset consists of 6511 restaurants and 591,980 dishes. Table 1 shows the summary statistics of the data.<sup>16</sup> There is a large variance in the number of dishes per restaurant (with a mean of 91 dishes and a standard deviation of 83). Many restaurants have few dishes (the mode is 38) but there is a large tail of restaurants with very long menus. Each dish is described with an average of 9.1 words, with each word containing on average 6.3 letters. Table 2 shows the geographical range of the data; some cities are underrepresented in the sample, and so in our main regression we add the city as a control variable.

	\$	\$\$	\$\$\$	\$\$\$\$
Total restaurants	2203	3311	844	153
Average number of dishes	114	88	50	48
Average description length (words)	7.6	10	9.8	9

Table 1. Summary statistics over the entire menu dataset. Description length is in words. Our dataset only includes food dishes, so the original menus would have been slightly longer including drinks.

City	# Restaurants
Boston	605
Chicago	480
Los Angeles	235
New York	2828
Philadelphia	624
San Francisco	1228
Washington, D.C.	511
Total	6511

<sup>16</sup>Middle priced restaurants in our dataset include chains (Ruby Tuesday's, T.G.I. Friday's, Cheesecake Factory, California Pizza Kitchen, Pasta Pomodoro, Red Lobster, Legal Sea Foods, Uno Chicago Grill) as well as individually owned restaurants like the White Dog Café in Philadelphia, Pizzeria Delfina in San Francisco, or Wu Liang Ye in New York.

Table 2. Number of restaurants per city in the data set.

### The hypotheses and coding

To develop a testable measure of the linguistic strategies corresponding to each aspect of distinction, we mainly measure the number of times words appear from *lexicons*, lists of words and phrases. Lexicons were drawn from the previous literature and also from an initial investigation of the menus, as described below. We also study summary properties like word length or sentence length.

For **natural versus traditional authenticity** we coded variables marking the two kinds of authenticity discussed in the Bourdieu-inspired food and menu literature.

1. Words and phrases related to the provenance of food (*natural, farmhouse, wild caught, grass fed, local, market, farmed, free range, heirloom*) derived from the previous literature,<sup>17</sup> as well as our initial investigation of the menus.
2. Words and phrases related to traditional authenticity drawn from Freedman and Jurafsky, as well as all mentions of family members (*old fashioned, traditional, family recipe, home style, mom, mother, auntie, grandpa, uncle, daddy's*)

For the concept of **plenty** we coded three variables indicative of Bourdieu's plenty and the related idea of the number of dishes or amount of choices the restaurant is offering the consumer:

1. the number of dishes on the menu (we controlled for the fact that menus with more dishes also have more words by adding as a control factor the number of total words on the menu)
2. phrases indicative of the extent of consumer choice (*choice of, choose, pick, specify, your own, your way, you like, any, add, or*)
3. phrases vaguely indicative of generous portions (*big, hearty, generous, huge, plenty, heaping*)

For **educational capital** we measured the log of the **average word length** in letters of all words describing all the dishes on the menu. Word length has a very strong inverse correlation with word frequency and a positive correlation with morphological complexity; long words are thus both rare and complex. Word length is thus the main measure of word complexity in measures of reading difficulty level like the Flesch-Kincaid readability test.<sup>18</sup> Word length is likely a less accurate proxy for morphological complexity on menus than it is in other genres, because of the multilingual nature of menus; for example words in Vietnamese are on average shorter than English, while words in Italian are longer. However, controlling for the restaurant type (which generally includes ethnicity) helps somewhat in dealing with this confound.

For **implicit status** we measured the number of adjectives in the menus. We ran the Stanford Part-of-Speech Tagger<sup>19</sup> on all menu descriptions, producing 5 million tagged words. Many menu descriptions, especially short ones, had incorrectly tagged words, presumably because the tagger was not trained on menus. The most frequent such error was caused by the fact that taggers often interpret long sequences of capitalized words standing alone as proper names; thus the participle “steamed” in the phrase “Steamed Little Neck clams” was incorrectly marked as a proper noun because of the neighboring (correctly tagged) proper noun “Little Neck”. However we found that the majority tag for most adjectives was in fact correct; cases like “steamed” were tagged correctly in the vast majority of instances (an instance of the *one sense per discourse*<sup>20</sup> rule). We therefore labeled as an adjective each instance of a word whose majority tag was as an adjective. Following usual practice, we eliminated all words that occurred fewer than 1 time per million words (i.e., had counts less than 5). We then hand-checked the remaining 1250 adjectives, eliminating remaining tagging errors. The result was 1065 hand-curated adjectives.

<sup>17</sup>Chahuneau et al., “Word Salad”; Freedman and Jurafsky, “Authenticity in America”; Sara Dickerman, “Eat Your Words: A Guide to Menu English,” *Slate* (April 29, 2003).

<sup>18</sup>Flesch-Kincaid: Rudolph Flesch. “A new readability yardstick.” *Journal of applied psychology* 32, no. 3 (1948): 221-233; J. Peter Kincaid, Robert P. Fishburne Jr, Richard L. Rogers, and Brad S. Chissom, *Derivation of new readability formulas (automated readability index, fog count and Flesch reading ease formula) for Navy enlisted personnel*. No. RBR-8-75. Naval Technical Training Command Millington TN Research Branch, 1975. Fog/smog: Robert Gunning, *The Technique of Clear Writing*, McGraw-Hill, 1952; G. Harry McLaughlin, “SMOG grading—a new readability formula,” *Journal of Reading* 12, no. 8 (1969): 639-646.

<sup>19</sup>Kristina Toutanova, Dan Klein, Christopher Manning, and Yoram Singer, “Feature-Rich Part-of-Speech Tagging with a Cyclic Dependency Network,” In *Proceedings of HLT-NAACL* (2003): 252-259.

<sup>20</sup>Gale, William A., Kenneth W. Church, and David Yarowsky. “One sense per discourse.” In *Proceedings of the workshop on Speech and Natural Language*, Association for Computational Linguistics (1992): 233-237; David Yarowsky, “Unsupervised word sense disambiguation rivaling supervised methods,” In *Proceedings of ACL*, ACL (1995): 189-196.

We also investigated some useful subsets of these 1065 adjectives. 164 *sensory adjectives* (*chunky, crispy, crunchy, doughy, fluffy, rich, smoky, tangy, zesty*) were selected by hand-coding the 1065, and checked to include all adjectives in Zwicky and Zwicky. 176 *participial* adjectives were all those among the 1065 that had participial morphology (*grilled, mixed, sliced, steamed, baked, smoked*). We also included two classes of positive sentiment adjectives, both drawn by extending the *extreme positive emotion* list that was carefully constructed by Larcker and Zakolyukina.<sup>21</sup> 55 *extreme positive sentiment* words were chosen by taking the 36 adjectives on their *extreme positive emotion* list that occurred more than 5 times in our corpus (more than 1 part per million) and that were not specifically food descriptive, and adding the 19 more adjectives that were synonyms of these. 14 *delicious* words were taken by separating the 4 food descriptive words (*delicious, delectable, scrumptious, luscious*) and adding synonyms (*tasty, gourmet, savory, mouthwatering, etc.*).<sup>22</sup>

**Control factors:** We added three control factors: the log of the total number of words describing all the dishes on the menu (used as an additional control for factors like the number of dishes), the city in which the restaurant was located, and the restaurant category, which consisted of a label from a set of 32 types of restaurants. These were constructed by grouping the 85 relevant Yelp restaurant types into the following 32 categories, based on choosing restaurants with similar cuisines and similar price ranges:

Pizza, Chinese, Italian, Steakhouses, American (new), Japanese, Mexican, French, American (traditional), Sandwiches, Cafes, Fast food, Thai, Indian, Other Asian, Diners, Seafood, Middle Eastern, Latin American, Bars, Bakeries, Spanish, Korean, Mediterranean, Barbeque, Other European, Vegetarian, Ethiopian, Soul food, Southern and Cajun, Greek, Asian fusion

Each restaurant was assigned to exactly one of these 32 categories; restaurants listed in Yelp with multiple classes were assigned to whichever of those classes occurred most frequently in the entire dataset.

## Experiment

We used a simple linear regression to test the relationship between variables and restaurant price status. The dependent variable is the number of dollar signs.

Our main independent variables are based on counts of words (a “bag of words” model since we ignore word order and other cues to grammatical structures). Word counts typically have a very long-tailed distribution so we apply the standard transformations to achieve a more Gaussian distribution. Thus each feature of a word count  $c$  was included as  $\log(1+c)$ , and length variables (length of dish product descriptions in words, and length of words in letters) were included as the log of the length (in letters or words).

Because all word count features drawn from a particular product description were collinear with the length of the product description (the more words in the description, the more chances for each type of word to appear), we next used a linear regression to remove effects of log length from each lexicon variable. We then included the residuals from this regression (rather than the raw counts) as variables to represent the effects of each lexicon of interest.

Linear regression was then used to predict the restaurant price (an integer from 1-4, representing \$, \$\$, \$\$\$, \$\$\$\$) from the variables of interest and the set of control variables, via the `lm` package in R.

Because the difference between the four categories (\$, \$\$, \$\$\$, \$\$\$\$) may not be linear, we also ran an ordinal logit using the `polr` function in R, predicting the four categories (\$, \$\$, \$\$\$, \$\$\$\$) as ordinal categories values from the independent variables. The results were the same as those from the linear regression, and so in the remainder of the paper we describe only the linear regression.

However, for the four different subtypes of adjectives considered below, we additionally ask whether some adjective classes were particularly associated with particular price classes (for example whether markers of traditional authenticity were especially associated with lower-middle priced (\$\$) restaurants). We therefore also ran multinomial logistic regressions (using `glm` in R) predicting a binary variable comparing the class of interest (in this case \$\$) against the other three classes

<sup>21</sup>David F. Larcker and Anastasia A. Zakolyukina, “Detecting deceptive discussions in conference calls,” *Journal of Accounting Research* 50, no. 2 (2012): 495-540.

<sup>22</sup>We return later to these implications of the modern meaning of the word “gourmet”.

(\$, \$\$\$, \$\$\$\$) and report on the role of the predictor in question as an independent variable, after controlling for all the same variables described above (restaurant category, number of dishes, number of words in different classes, etc.).

## Results

We describe the results of each aspect of distinction separately; full regressions are shown in the appendix.

Restaurant price is significantly associated with the language of natural authenticity ( $p < 2 \times 10^{-16}$ ); expensive restaurants have descriptions like "Local Albacore Tuna Nicoise: *summer* beans, *heirloom* tomatoes, yellow tomato aioli, soft boiled *farm* egg" or "Bison Burger: 8 oz. Blue Star Farms, *grass fed & pasture raised*".

Traditional authenticity is associated instead with cheaper restaurants ( $p = .000716$ ), especially lower-middle-priced (\$\$) restaurants which offer dishes like "Homemade Tiramisu: family recipe", "Old Fashioned Beef Stew", or "Annie's Famous Pot Roast: Homemade, just like Mom's." Figure 1 shows the raw counts and confidence intervals.

To test the hypothesis (suggested from Figure 1) that it was lower-middle priced (\$\$) restaurants that make extra use of traditional authenticity, we also ran a logistic regression predicting class (\$\$ versus \$, \$\$\$, \$\$\$\$) after controlling for restaurant category, number of dishes, and number of words. The framing of traditional authenticity indeed was more likely to be used in the menus of lower middle priced restaurants ( $p = 0.000303$ ).

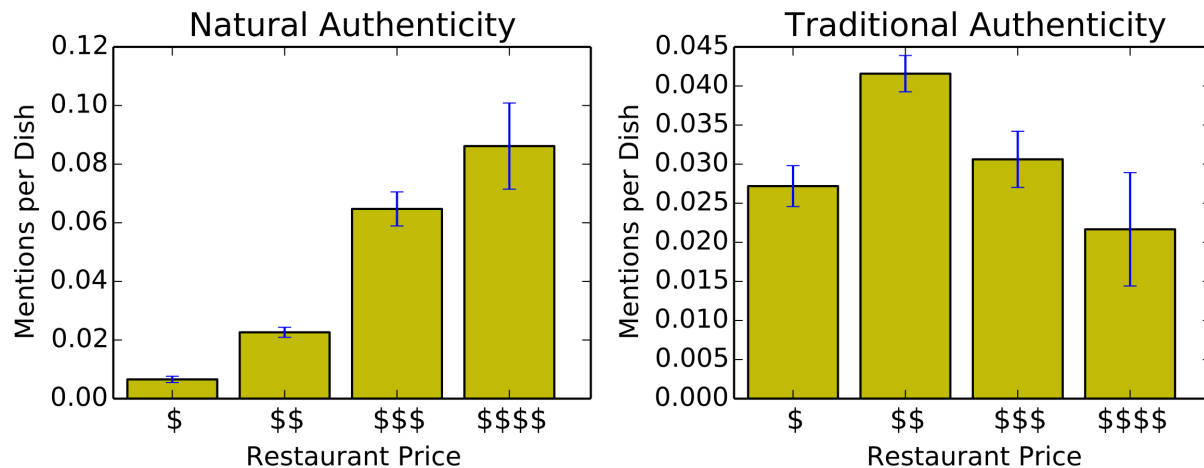


Figure 1. Means and 95% confidence intervals for raw values of natural versus traditional authenticity by restaurant price status. The values are counts per average dish description. Thus for example on \$\$\$ menus, mentions of natural authenticity occurred on average 6 times in every 100 dish descriptions.

Educational capital was also associated with restaurant status. More expensive restaurants were indeed more likely ( $p < 2 \times 10^{-16}$ ) to use longer words like *accompaniments*, *complements*, *magnificent*, *inspiration*, *exquisitely*; cheaper restaurants used shorter forms like *sides* instead of *accompaniments* or *complements*. Figure 2 shows raw means and standard errors:

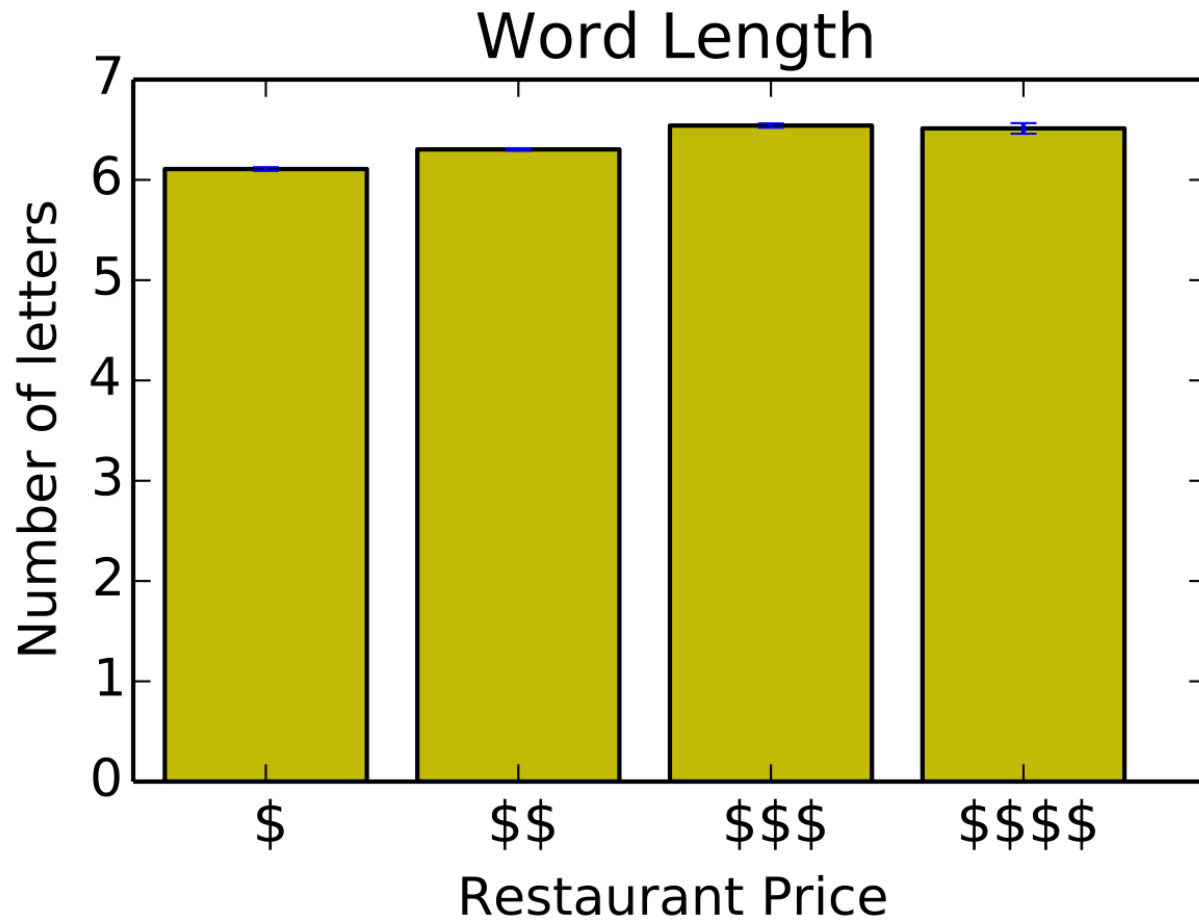


Figure 2. Means and 95% confidence intervals for raw values of word length in letters by restaurant price status.

To see the association of foreign words with expensive restaurants, we examined the words most likely to occur in the \$\$\$\$ class, which we defined as those words with the highest log likelihood ratio between the counts in the \$\$\$\$ class and the summed counts in the other three classes, using the “weighted log-odds-ratio, informative Dirichlet prior” method.<sup>23</sup> We found that 28 of the 100 words most associated with the expensive restaurants were foreign, including 9 French words (*les, de, le, fois, gras, crème, mousse, tarte, pommes*), 6 Italian words (*e, con, risotto, pancetta, burrata, polenta, parmigiano*), and 13 Japanese words (*tempura, uni, wagyu, sushi, yuzu, sashimi, miso, shabu, kobe, ponzu, wasabi, ninja, and sake*). The fact that the French and Italian vocabularies are full of function words (the short grammatical words like *de* or *con*) suggests that at least some of these product descriptions are written largely or even entirely in French or Italian. In summary, more expensive restaurants had longer words and more high-status foreign words.

As expected cheaper restaurants have more dishes on the menu ( $p < 5.7 \times 10^{-15}$ ) and are more likely to use the framing of choice and plenty. They emphasized generous portion sizes, making use of words like *generous, hearty, and big* ( $p = .0023$ ) (“Baked Meat Lasagna: Made with our *hearty* meat sauce,” “Tuna Supreme: A *generous* scoop of tuna”). Cheaper restaurants are more likely to emphasize the extent of diner’s choice with descriptions like “baby lamb chops, grilled *to your liking*,” “PEI mussels: *choose your style*”) ( $p < 2 \times 10^{-16}$ ).

<sup>23</sup>Burt L. Monroe, Michael P. Colaresi, and Kevin M. Quinn, “Fightin’ words: Lexical feature selection and evaluation for identifying the content of political conflict,” *Political Analysis* 16, no. 4 (2008): 372-403.



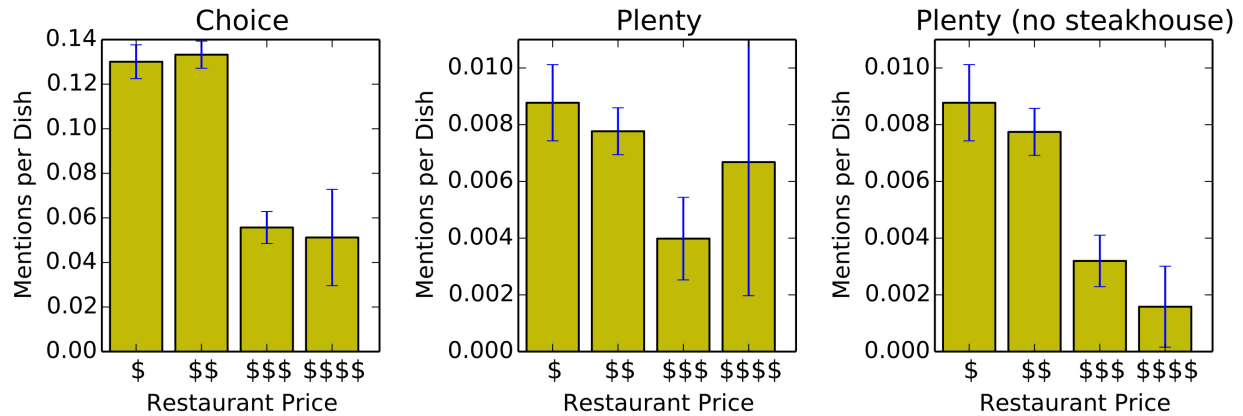


Figure 3. Means and 95% confidence intervals for raw values of choice and plenty by restaurant price status. 3b shows values of plenty for all restaurants; note the very large variance for the \$\$\$\$ restaurants; 3c shows values for the restaurants after removing all steakhouses.

Figure 3 shows raw means and variances. Note from Figure 3b that while plenty was significantly associated with cheaper restaurants, there is still enormous variance in the most expensive \$\$\$\$ class. We examined the uses of plenty framing in \$\$\$\$ restaurants and found that most of these were steakhouses; here's a sample usage from a Chicago steakhouse:

Big Shoulders: and you thought our petite porterhouse was big! this is a full forty ounces! that's eight ounces short of three pounds! are you up to the challenge?

Figure 3c shows the vastly reduced variance in a smaller version of the dataset that removed the 81 steakhouses. To further investigate this suggestion of an association between steakhouses and plenty, we also ran a linear regression predicting the log number of mentions of words suggesting plenty from the restaurant category. Seven classes of restaurants were more likely to use these framings: steakhouses ( $p = 2.30 \times 10^{-7}$ ), traditional American food ( $p = 4.17 \times 10^{-9}$ ), fast food ( $p < 2 \times 10^{-16}$ ), sandwiches ( $p = 6.28 \times 10^{-11}$ ), bars ( $p = 0.00048$ ), pizza ( $p = 0.0044$ ) and barbecue ( $p = 0.00019$ ). Our findings thus confirm the association of plenty and choice with cheaper restaurants; we return in the discussion section to the implications of the association of plenty with steakhouses and other types.

Finally, we examined the role of adjectives. Conforming to our hypothesis, adjectives in general were negatively associated with price; cheaper restaurants used more adjectives than expensive ones ( $p < 6.1 \times 10^{-5}$ ). Figure 4 shows the raw means and variances for the four specific sub-classes of adjectives we consider.

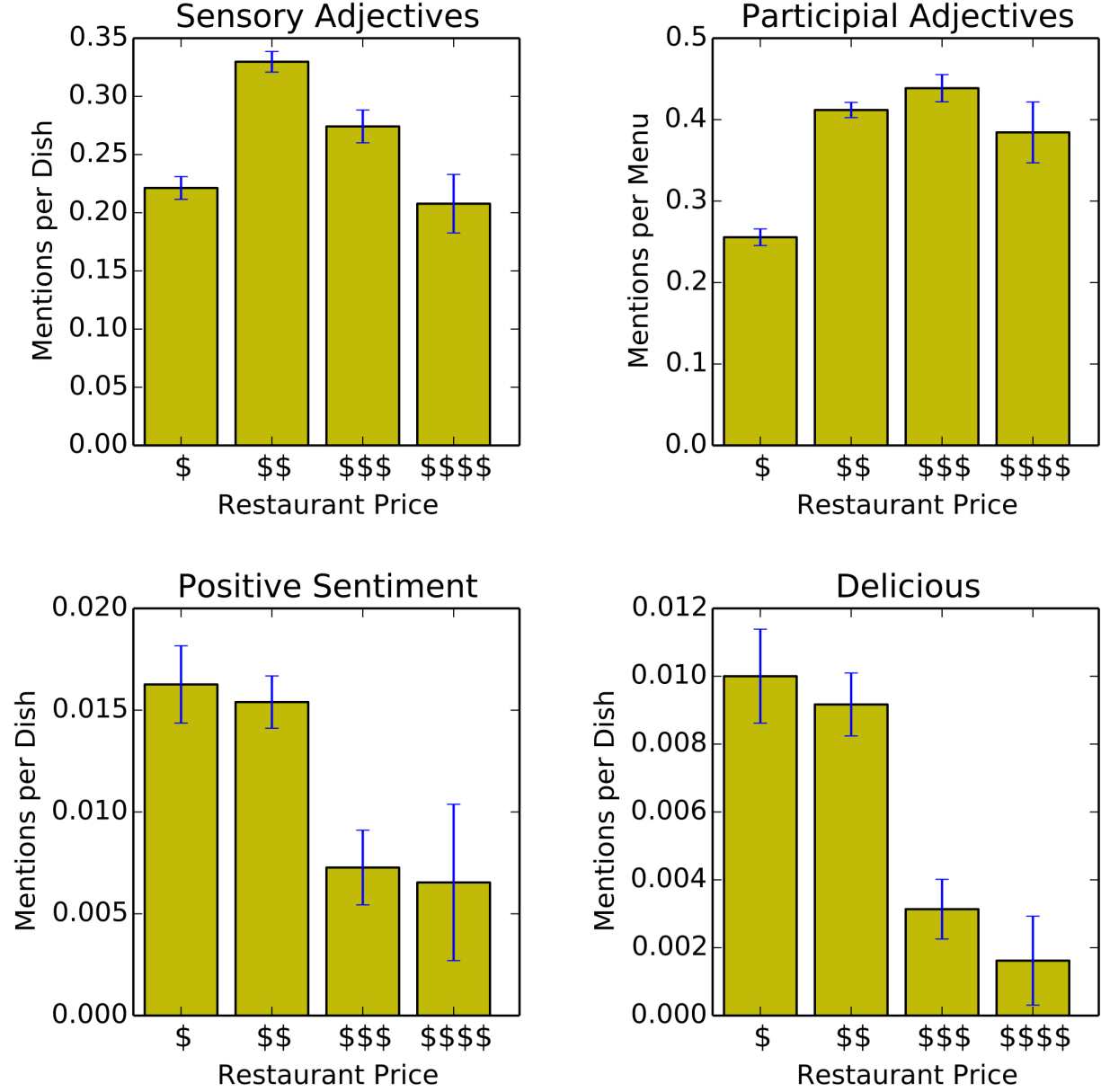


Figure 4. Means and 95% confidence intervals for raw values of four classes of adjectives by restaurant price status. Sensory adjectives and participles are more associated with middle-priced restaurants, while adjectives of positive sentiment (either related to food or not) are associated with the cheapest restaurants.

As Figure 4 suggests, while adjectival protestations of quality are used less on expensive restaurants, different classes of adjectives are associated with different groups of cheaper restaurants. Sensory adjectives are strongly associated with lower-middle-priced restaurants ( $p = 0.000182$ ) in a logistic regression comparing \$\$ with \$,\$\$\$,\$\$\$\$ after controlling for category, number of dishes, length and the other adjectives). We see examples like "*Crisp Golden Brown Belgian waffle with Fresh Fruit*". Participles are similarly associated with lower-middle-priced (\$\$) restaurants (same regression,  $p = 1.66 \times 10^{-8}$ ). Both positive sentimental adjectives (*excellent, great, wonderful*,  $p = 9.09 \times 10^{-8}$ ) and positive food adjectives (*delicious, tasty, gourmet*,  $p = 1.76 \times 10^{-7}$ ) are associated with lowest priced (\$) restaurants in a logistic regression comparing \$ with \$,\$\$\$,\$\$\$\$ after controlling for category, number of dishes, length and the other adjectives.

## Studying menus over time

Our study suggests a number of ways that modern menus reflect Bourdieu's ideas of distinction. But it is important to understand to what extent these factors are long-lasting trends that describe attitudes toward food culture in the United States, and to what extent they solely characterize recent food culture in 2011, the date the menus were collected. The fact that Bourdieu's own analyses date from data he collected in the 1960s, half a century ago in France suggests that these trends have persisted for a while. Have cheaper restaurants always emphasized traditional authenticity, and expensive ones natural authenticity?

To answer this question we turn to a second dataset: the New York Public Library's Buttolph Collection, which contains over 17,000 digitized and transcribed menus from 1852-2015.<sup>24</sup> The menus come from a wide variety of locales, but the majority of those whose source is labeled come from New York City restaurants or from steamships. The collection is strongest around the turn of the last century, so we extracted all menus from 1892-1921. Because the dataset has enormous variation across time and space and the metadata is often lacking, our analysis here is necessarily preliminary, leaving for future investigation many questions like establishing the correct control variables (restaurant category, exact location) or of designing a proper sample. Nevertheless, we chose to do a preliminary analysis, controlling for at least some variables. For example because the dataset contains enormous numbers of duplicate menus from the same restaurant on different days (210 menus in the year 1914 alone come just from different lunches and dinners at the Waldorf Astoria Hotel), we eliminated all cases of identical dishes from the same restaurant in the same year, resulting in 2858 distinct menus for this period. We defined a simplified 2-point pricing scale (cheap versus expensive), first using the median price of the dishes on a menu to assign a cost score to each menu, then computing an annual median of the menu costs for each year in the data, and labeling a menu in a given year "expensive" if its cost score was above the median for that year, and "cheap" otherwise.

We then asked which of the aspects of distinction discussed above characterized menus from a century ago. Multinomial logistic regression was used, with the price of a menu (cheap versus expensive) as the dependent variable, and each of the textual indicators of distinction discussed above as independent variables. Only five variables were significantly associated with price. More expensive menus had longer words ( $p < 2 \times 10^{-16}$ ), and cheaper menus displayed more linguistic signs of traditional authenticity ( $p = 3.45 \times 10^{-7}$ ) and choice ( $p = 1.02 \times 10^{-6}$ ), and used more adjectives ( $p = 1.07 \times 10^{-9}$ ). One control variable was also significant: cheaper menus used more words overall ( $p = 0.000474$ ). None of the other variables were significantly different between the cheap and expensive restaurants (although there were large differences in the presence of French vocabulary, which we return to below). Figure 6 shows the raw means and variances for these variables.

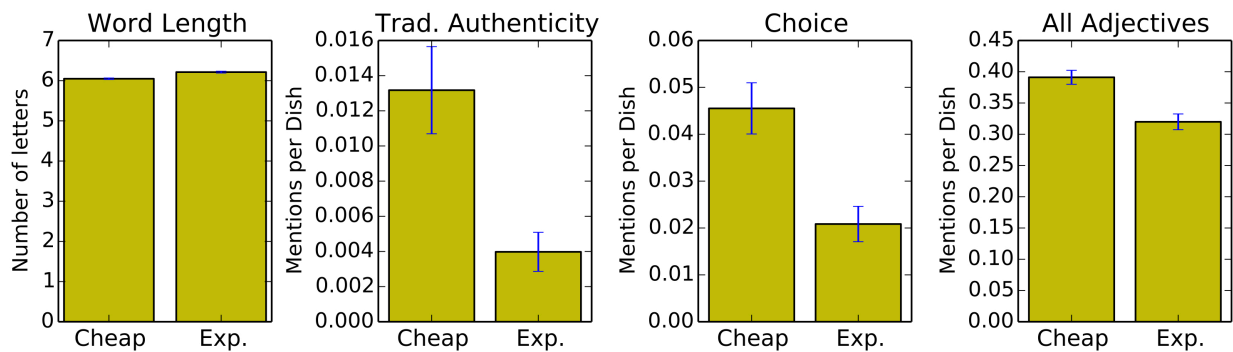


Figure 5. Means and 95% confidence intervals for raw values of four variables in historical menus from 1892-1921 from the New York Public Library Buttolph Collection. These four variables all pattern exactly as on modern menus. Other variables signaling distinction (number of items, plenty, sensory adjectives, natural authenticity, sentiment adjectives, delicious words) show no interaction with restaurant price in these menus.

Early examples of traditional authenticity on inexpensive menus look very much like modern menu items, with examples like "homemade sausage with hot slaw" from the 1899 menu of the Hotel Baltimore, or "old fashioned corn bread with

<sup>24</sup>Michael Lascarides and Ben Vershbow, "What's on the menu? Crowdsourcing at the New York Public Library," In *Crowdsourcing Our Cultural Heritage*, ed. Mia Ridge. 2014.

maple syrup” from the 1900 menu of Mandel’s Tea Rooms. Customer choice also appears on cheaper menus in the previous century, such as Haan’s advertising “terrapin any style” in 1900 (turtle was very popular at that time), Hazeltine’s 1914 “toasted bread any kind”, or the 1913 menu of Smith’s Restaurant which allows the customer to “choose one dish from each column.”

The results of our investigation of menus from a century ago, while certainly preliminary, suggests that at least some of the aspects of Bourdieu’s distinction have been visible on restaurant menus for well over a century. The fact that lower-priced restaurants emphasized traditional authenticity suggests that the framing of inexpensive food as old-fashioned or homey has itself a long history in the US. The fact that expensive restaurants use longer words points to the early role of educational capital in marking status.

We did not find evidence that expensive restaurants were more likely to mention farms or pastures. Restaurants of that period certainly did mention farms or the way the food was raised (“Deerfoot Farm sausages”, “buttermilk from Darling Farm”, “special raised turkey”), but these were not more likely to be expensive restaurants. What words were then most likely to occur in the most expensive restaurants? We examined the words with the highest log likelihood ratio between the counts in the restaurants in the top quartile of priced class and the summed counts for all other restaurant price classes, again using the “weighted log-odds-ratio, informative Dirichlet prior” method of Monroe, Colaresi and Quinn (2008). Of the 20 words ranked as most characteristic of the most expensive restaurants, 14 were French words (*creme, salade, de, la, a, au*, etc.). The remainder were mainly expensive products (*lobster, squab, champagne*, and, perhaps surprisingly to the modern reader, *chicken*). The high prevalence of French words on expensive menus reflects the well-known role of French language and cuisine as a signifier of high status in American cuisine of the 19th and 20th centuries.

The fact that the expensive restaurants also use fewer words and fewer adjectives suggests that the use of implicit or “modest advertising” to mark high status dates back at least 100 years. Individual words, however, have changed their association with status over time, in the kind of trickle-down of status suggested by Veblen and Simmel.<sup>25</sup> In the historical corpus, for example, *gourmet* is associated with more expensive restaurants, like the “braised lamb gourmet” from the 1917 Waldorf Astoria or the “tomato gourmet salad” from the 1920 Hotel Manhattan. In modern menus, however the word *gourmet* is associated only with cheaper restaurants, and you can order a “gourmet large pepperoni” from any number of pizza delivery restaurants in the 2011 dataset. While in many uses *gourmet* surely retains some of its earlier sense of epicurean expertise, on menus the word has undergone semantic bleaching, leaving it mainly as a vague protestation of quality.

## Discussion

We investigated a large dataset of restaurant menus coded with textual, quantitative, and metadata information to understand how words on menus subtly reflect aspects of our food culture. Our results suggest that the words used in menu descriptions of American restaurants reflect the many aspects of Bourdieu’s *distinction*. Expensive restaurants use the language of natural authenticity, focusing on the provenance of their food, while cheaper restaurants focus on traditional authenticity, highlighting old-fashioned American dishes and moms and grandmas. The link between the natural and pastoral with expensive menus is also consistent with a wide variety of previous research, and is presumably linked with the huge changes in American food that occurred in the 1960s and 1970s.<sup>26</sup> Researchers have noted a contemporaneous increase in emphasis on the pastoral in the related field of wine reviewing, and suggested that the association of the pastoral genre with the urban upper class could also be at least in part a response to the increasing role of technology and related conflicts of modernity.<sup>27</sup>

The use of these two metaphors for authenticity acts as a way of targeting consumers, using values that appeal differently to the two groups; food origin and purity for higher socio-economic status consumers, and family and tradition for lower

<sup>25</sup>Thorstein Veblen, *The Theory of the Leisure Class: An Economic Study of Institutions*, 1899; Georg Simmel, “Fashion,” *International Quarterly* 10 (1904): 130-155.

<sup>26</sup>Harvey A. Levenstein, *Revolution at the table: the transformation of the American diet* (University of California Press, 2003); Michael Symons, “Grandmas to gourmets: The revolution of 1963,” *Food, Culture & Society* 9, no. 2 (2006): 179-200.

<sup>27</sup>Sean Shesgreen, “Wet Dogs and Gushing Oranges: Winespeak for a New Millennium,” *The Chronicle of Higher Education*. March 7, 2003; Freedman and Jurafsky, “Authenticity in America”.

socio-economic status consumers.<sup>28</sup> Our work shows that authenticity is not a monolithic concept: it can be used to mean different things to different consumers, and that firms use these different kinds of authenticity in a coherent way for product differentiation. Furthermore, at least one aspect of this framing, the association of tradition with cheaper restaurants, dates back well over a century.

Modern expensive restaurants are more likely to use longer words and to use foreign words from three high-status foreign languages: French, Italian, and Japanese. How do complex and foreign words acquire their linguistic association with high status? French of course has long been associated with high status in cuisine; Italian and Japanese are more recent high-status culinary languages. The use of more complex words is a sign of educational capital, which has long been associated with high status. Complex words are also associated with more formal genres<sup>29</sup> and more formal or ceremonial language is also associated with high status or luxury. It's possible as well that these complex words are an attempt to appeal to the reader's ego; a reader who can understand these complex or foreign words is implicit being flattered at their ability to understand the "code". Presumably, a more complex menu also requires more educated or trained wait staff to discuss the menu with customers. This association of complex words, and especially French words, with high status was visible also in menus from a century earlier.

Cheaper restaurants focus on *plenty* by emphasizing the size of portions (*generous*, *Texas-sized*) and the amount of choices they offer in number of dishes and options within those dishes. This framing is most associated with steakhouses, traditional American food, fast food, pizza, sandwiches, barbecue, and bars. Except for steakhouses, these are not fancy restaurants, but everyday places that seem to cater to the eater more concerned with value. Furthermore, these are all restaurants that serve American food, focusing on American ethnic foods (sandwiches, hamburgers, and barbecue); thus Chinese restaurants, while prevalent in the cheap (\$) categories, do not seem to make use of the metaphor of plenty. Steakhouses are the exception to the generally low prices of restaurants using this metaphor. Although expensive in price, steakhouses seem to draw on this same framing of the all-American working class, suggesting we should expect to see other aspects of working class framing in steakhouses. The fact that only part of this framing of plenty (the emphasis on choice, but not the emphasis on large portions) was present a century earlier suggests a change in this aspect of distinction.

Finally, expensive restaurants generally use more implicit language, using fewer of the many kinds of adjectives used by cheaper restaurants, and this use of fewer words and adjectives was also present in menus a hundred years ago. H. Paul Grice's model of cooperative language behavior may explain how the avoidance of such modifiers sends a signal about quality. Grice proposed that language users automatically obey certain communicative maxims.<sup>30</sup> A rational communicator of the type described by Grice would only mention that food is *crisp* or *fresh* to fulfill a communicative goal like convincing the reader, which only makes sense if the reader doesn't already think the food is crisp and fresh. A high status restaurant, however, wants freshness and crispness to be already presumed, and therefore the crispness should go unmentioned. No expensive restaurant would use a description like this one from a cheap restaurant: "a *flavorful*, *colorful*, and *delicious* salad mixture of *crispy* bacon bits". For an expensive restaurant that the food is flavorful and delicious should go without saying.

To see these Gricean inferences in a qualitative way, we examined the foods most commonly associated with one of these adjectives, "real". Exactly which foods a menu claim to be "real" depends on the price. The least expensive (\$) restaurants are most likely to promise "real whipped cream", "real mashed potatoes", and "real bacon". In slightly more expensive (\$\$) restaurants, "real" is used mainly to describe "real crab" and "real maple syrup". By contrast, "real" is barely used at all for more expensive (\$\$\$ and \$\$\$\$) restaurants. For a pricy restaurant to call its crab "real" would be to suggest that its realness might be in question and has to be defended.

The avoidance of over-explicit adjectives by high status restaurant is also consonant with the game-theoretic models described above,<sup>31</sup> and with consumer studies that find that consumers perceive excessive ads as overcompensating for

<sup>28</sup>Alana Conner Snibbe and Hazel Rose Markus, "You can't always get what you want: educational attainment, agency, and choice," *Journal of personality and social psychology* 88, no. 4 (2005): 703-720; Nicole M. Stephens, Stephanie A. Fryberg, and Hazel Rose Markus, "When Choice Does Not Equal Freedom: A Sociocultural Analysis of Agency in Working-Class American Contexts," *Social Psychological and Personality Science* 2, no. 1 (2011): 33-41.

<sup>29</sup>Douglas Biber, *Dimensions of Register Variation: A cross-linguistic comparison* (Cambridge University Press, 1995); Julian Brooke, Tong Wang, and Graeme Hirst, "Automatic acquisition of lexical formality," In *Proceedings of the 23rd International Conference on Computational Linguistics*, Association for Computational Linguistics (2010): 90-98.

<sup>30</sup>H.P. Grice, *Studies in the Way of Words* (Harvard University Press, 1989).

<sup>31</sup>Orzach and Tauman, "Modest advertising signals strength"; Feltovich et.al., "Too cool for school?"

problems in a low-quality product.<sup>32</sup> The visual counterpart of minimal language is the use of white space in spare ads, which both consumers and creative directors associated with prestige and quality, based on a link between white space and the mid-century minimalist movement in art and the “less is more” movement in architecture, all of which associated spare, clean, minimal designs with prestige and the upper class in North America.<sup>33</sup> Finally, this use of overly explicit descriptions by less expensive restaurants might also be modeled as a kind of overcompensation, in which a group that is anxious about its status overcompensates in the cues for that status; this kind of overcompensation is common both in linguistics, in the hypercorrection seen in speakers of non-standard styles,<sup>34</sup> or the overcompensation seen in other social categories like masculinity.<sup>35</sup>

While our results are generally consistent with prior literature, there are some differences. Our finding that inexpensive restaurants are more likely to emphasize consumer choice as compared with expensive restaurants seems inconsistent with work finding that European-Americans of higher socio-economic status (SES) emphasize personal choice more than those of lower SES.<sup>36</sup> It is possible that choice is simply more implicit in high-SES restaurants, which may be more willing to make special versions of dishes (for example omitting particular ingredients), have waiters discuss choices orally, or put choices on a chalkboard without highlighting this fact on the written menu. Investigating this paradox is an important direction for future research.

Our study has a number of limitations, such as its focus solely on major metropolitan areas and its focus on firms solely in the United States. Strauss’s (2005) comparative study of the language of television advertising shows that generic terms like *oishii* or *umai* (‘delicious’ or ‘tasty’) are very common in Japanese advertising, but generic terms like ‘delicious’ are less common in American or Korean advertising, which are instead more likely to discuss specific positive qualities that cause the food to be delicious. Thus the finding that vague positive words like *delicious* or *tasty* are associated with lower priced restaurants or dishes is likely specific to marketing in the United States context. Further cross-cultural research is clearly called for. Understanding the role of the socio-economic status of the consumer, and also how these restaurant meals fit into larger patterns of food purchase and consumption, also remain important directions.

Despite these caveats, our results highlight the important role that investigating the “linguistics of the everyday” should play for our understanding of culture. Quotidian aspects of life are useful windows onto culture, not just because our attitudes toward daily life reflect our implicit beliefs about identity and socio-economic class, but also because they may come pre-annotated with economic variables, as does the language of restaurant menus or food reviews. Computational techniques can thus be key in helping explore aspects of culture and society.

## Appendix

Category	#	%	Category	#	%
Pizza	730	11.20%	Middle Eastern	121	1.90%
Chinese	615	9.40%	Diners	105	1.60%
Italian	592	9.10%	Seafood	96	1.50%
American (new)	494	7.60%	Steakhouses	81	1.20%
American (traditional)	473	7.30%	Mediterranean	59	0.90%
Japanese	462	7.10%	Spanish	54	0.80%
Mexican	397	6.10%	Korean	52	0.80%
Bars	276	4.20%	Bakeries	52	0.80%

<sup>32</sup>Amna Kirmani, “The effect of perceived advertising costs on brand perceptions,” *Journal of Consumer Research* 17, no. 2 (1990): 160-171; Amna Kirmani, “Advertising repetition as a signal of quality: If it’s advertised so much, something must be wrong,” *Journal of Advertising* 26, no. 3 (1997): 77-86; Pamela M. Homer, “Ad size as an indicator of perceived advertising costs and effort: The effects on memory and perceptions,” *Journal of Advertising* 24, no. 4 (1995): 1-12.

<sup>33</sup>J. W. Pracejus, Olsen, G. D., & O’Guinn, T. C., “How nothing became something: white space, rhetoric, history, and meaning,” *Journal of Consumer Research* 33, no. 1 (2006): 82-9; John W. Pracejus, Thomas C. O’Guinn, and G. Douglas Olsen, “When white space is more than”burning money: Economic signaling meets visual commercial rhetoric,” *International Journal of Research in Marketing* 30, no. 3 (2013): 211-218.

<sup>34</sup>William Labov, “Hypercorrection by the lower middle class as a factor in linguistic change,” In *Sociolinguistics*, ed. William Bright. Mouton, 1966.

<sup>35</sup>Robb Willer, Christabel L. Rogalin, Bridget Conlon, and Michael T. Wojnowicz, “Overdoing gender: A test of the masculine overcompensation thesis,” *American Journal of Sociology* 118, no. 4 (2013): 980-1022.

<sup>36</sup>Conner Snibbe et.al., “You can’t always get what you want”; and Stephens et al., “When Choice Does Not Equal Freedom”.

Category	#	%	Category	#	%
Sandwiches	226	3.50%	Asian fusion	46	0.70%
Thai	220	3.40%	Barbeque	36	0.60%
Other	200	3.10%	Other European	34	0.50%
Fast food	199	3.10%	Vegetarian	32	0.50%
French	197	3.00%	Ethiopian	21	0.30%
Indian	178	2.70%	Greek	17	0.30%
Latin American	165	2.50%	Southern	13	0.20%
Coffee & Tea	132	2.00%	Soul food	13	0.20%
Other Asian	123	1.90%			
			Total	6511	

Table A1. Broad restaurant categories for all restaurants in the dataset. Categories were created by hand-grouping Yelp categories by cuisine and price similarity.

	Estimate	StdError	z_value	Pr(> z )	
(Intercept)	-2.856242	0.329042	-8.68	<2e-16	***
citychicago	-0.043169	0.033967	-1.271	0.203809	
cityla	0.061897	0.042524	1.456	0.145562	
citynyc	0.166924	0.02487	6.712	2.08E-11	***
cityphiladelphia	0.018239	0.031218	0.584	0.559081	
citysf	-0.047399	0.027739	-1.709	0.087544	.
citywashington	0.102401	0.033002	3.103	0.001925	**
categoryamerican_(traditional)	-0.226379	0.036237	-6.247	4.44E-10	***
categoryasian_fusion	0.155454	0.084581	1.838	0.066119	.
categorybakeries	-0.53602	0.080458	-6.662	2.92E-11	***
categorybarbeque	0.122788	0.095623	1.284	0.199156	
categorybars	-0.210265	0.041823	-5.028	5.10E-07	***
categorychinese	-0.266647	0.040417	-6.597	4.52E-11	***
categorycoffee_&_tea	-0.573856	0.054413	-10.546	<2e-16	***
categorydiners	-0.290413	0.061352	-4.734	2.25E-06	***
categoryethiopian	-0.143439	0.122629	-1.17	0.242165	
categoryfast_food	-0.549233	0.047937	-11.457	<2e-16	***
categoryfrench	0.331963	0.046377	7.158	9.10E-13	***
categorygreek	-0.287144	0.134892	-2.129	0.033317	*
categoryindian	-0.319241	0.049826	-6.407	1.59E-10	***
categoryitalian	0.003328	0.034616	0.096	0.923404	
categorykorean	-0.070347	0.081356	-0.865	0.387249	
categorylatin_american	-0.332607	0.050576	-6.576	5.20E-11	***
categorymediterranean	-0.333472	0.075645	-4.408	1.06E-05	***
categorymexican	-0.489798	0.038614	-12.684	<2e-16	***
categorymiddle_eastern	-0.426998	0.056383	-7.573	4.15E-14	***
categoryother	-0.203001	0.046395	-4.375	1.23E-05	***
categoryotherasian	-0.372029	0.057198	-6.504	8.40E-11	***
categoryothereuropean	-0.289007	0.097082	-2.977	0.002922	**
categorypizza	-0.64579	0.034175	-18.897	<2e-16	***
categorysandwiches	-0.85918	0.044656	-19.24	<2e-16	***
categoryseafood	0.301452	0.061283	4.919	8.91E-07	***
categorysoul_food	-0.036079	0.154407	-0.234	0.815257	
categorysouthern	-0.1077	0.153362	-0.702	0.482544	
categoryspanish	0.110111	0.078953	1.395	0.163175	
categorysteakhouses	0.928564	0.065865	14.098	<2e-16	***

	Estimate	StdError	z_value	Pr(> z )	
categorysushi_bars	0.205773	0.038764	5.308	1.14E-07	***
categorythai	-0.193253	0.04799	-4.027	5.71E-05	***
categoryvegetarian	-0.198101	0.099919	-1.983	0.047453	*
logwordlen	2.314061	0.160053	14.458	<2e-16	***
loglength	0.274683	0.016534	16.613	<2e-16	***
lognumitems	-0.082184	0.010498	-7.828	5.74E-15	***
adjectives	-0.228804	0.052673	-4.344	1.42E-05	***
plenty	-0.931954	0.305653	-3.049	0.002305	**
consumer choice	-0.595383	0.06144	-9.691	<2e-16	***
natural authenticity	2.308947	0.159477	14.478	<2e-16	***
traditional authenticity	-0.45395	0.134103	-3.385	0.000716	***

Table A2. Coefficients for the linear regression predicting restaurant price status (an integer 1-4 corresponding to \$, \$\$, \$\$\$, \$\$\$\$) from control variables and variables of interest, using the lm package in R. The adjusted R-squared was 0.4458. All variables from the count lexicons (adjectives, plenty, etc.) are residuals after regressing out the effects of description length in words (loglength).

Sensory Adjectives/Adverbs: airy, aromatic, astringent, beefy, bitter, bittersweet, blazing, bloomy, bold, bright, briny, brisk, burnt, buttery, c  
Positive Sentiment: amazing, appealing, awesome, beautiful, best, better, dazzling, delightful, divine, dynamite, excellent, exceptional, excit  
Positive Food Sentiment: appetizing, delectable, delicious, flavorful, gourmet, luscious, mouthwatering, savory, scrumptious, tastiest, tasty,  
Plenty: big, bigger, biggest, bottomless, bountiful, colossal, endless, enormous, generous, generously, gigantic, ginormous, heaped, heaping  
Choice: choice, choose, any, add, or, specify, substitutions, specifications, options, pick, your way, your own, your liking, your style, your fa  
Traditional Authenticity: home\*, traditional\*, timeless, family recipe, all american, our founder, old fashioned, old school, american favorit

Table A3. Lexicons